EXPERIMENT IN SELF-INSTRUCTION IN GENERAL BIOLOGY*

EDWARD J. KORMONDY AND ELLIS L. VAN ATTA

Department of Biology and Department of Psychology, Oberlin College, Oberlin, Ohio

A continuing interest in discovering ways of increasing instructional efficiency and providing suitable materials for independent study prompted the development and testing of "programmed materials" for use in self-instruction devices or "teaching machines" in introductory zoology, psychology, and music theory at Oberlin College. The rationale of the teaching machine and its applicability in independent study have been discussed by Skinner (1958); reviews of the present state of the field occur in Galanter (1959) and Silverman (1960); a source book has been prepared by Lumsdaine and Glaser (1960).

In sum, the available evidence is persuasive that teaching machine techniques increase learning efficiency, a desideratum achieved because the materials are designed on the basis of certain principles in the psychology of learning. Programmed material: (1) allows each student to proceed at his own rate; (2) requires him to participate actively in the learning process by calling on him to supply answers continually during a period of work with the programmed materials; (3) furnishes him with immediate knowledge about the correctness or incorrectness of his answer; (4) gives him a way of knowing, on his own, when he has thoroughly learned the material; and (5) provides a more thorough and systematic coverage of what is to be studied, instead of leaving up to the student (who may do it well or badly) the selection and organization of materials to be studied. In short, teaching machine programs, properly designed, possess the characteristics which are required of materials suitable for independent study.

With the use of self-instructional techniques designed to teach basic information, the student can acquire a body of fundamental knowledge on his own with little or no direct supervision from his instructor. The teacher can thus be freed from part of the obligation first to present basic information and then to check up on the student's learning. Class meetings can be used for activities that require the presence and direction of the teacher, such as: (1) class discussion of the controversial, ambiguous, and evaluative aspects of the field, in which the development of judgment is essential; (2) application of techniques of interpreting, criticizing, and developing experiments, ideas, arguments, or points of view; and (3) the conveying, by example, of a sense of how a trained scholar or investigator goes about his work.

The subject matter of the biology program comprised a review of mitosis and an introduction to meiosis and basic genetics. This section of the course contains basic information suitable for programming and complex and technical aspects fundamental to the mastery of a significant portion of the content of the course. The programmed materials were first given during the fall semester of 1959–60 to 184 Oberlin College students enrolled in an introductory course in zoology, and, in collaboration with Dr. William K. Stephenson, of the Department of Biology of Earlham College, to 161 beginning biology students at Earlham College. A revised version of the program was given during the fall semester of 1960–61 to 213 Oberlin College students.

The program was written according to the procedures of B. F. Skinner of Harvard University. This technique of programming results in what has been

*The "teaching machine" project began in June 1959 under a grant from the Ford Foundation.

called a "linear" program; that is, the subject matter to be programmed is first broken down into major component sub-topics, and these are arranged in some sort of logical sequence. Each component part of the program is then further broken down into individual items, or questions, which are arranged in an orderly sequence designed to lead by very small steps toward mastery of that particular sub-topic. Each item is in the form of a "completion" question, in which one or more key words are omitted and the student is required to "fill in the blanks" with the word or words which will best complete the item. The answer to each item is not available to the student at the time he makes his response, but is provided immediately following his response. Thus, in a typical sequence, the student reads an item, records his response to that item on an answer sheet, exposes the answer provided by the program as correct and checks his response against it, and then moves on to the next item where the process is started over again. In this manner, he works his way through an entire lesson, being informed at each step as to whether he is correct or incorrect.

The linear program used in the present study should be distinguished from the "instrinsic" program used by Crowder (1960). Intrinsic programming results in a series of multiple-choice items. If the student selects the correct alternative on any item, he proceeds to the next item. If he selects an incorrect alternative, he is informed why the choice is incorrect and is then directed to re-read the item and to select another alternative. In its structure, the intrinsic program is similar to the self-scoring tests originated by Pressey (1926) and to the type of program with which he is currently most closely identified (Pressey, 1960).

MATERIALS AND EXPERIMENTAL DESIGN

In the absence of available teaching machines in the fall of 1959 the programmed materials were arranged in a format referred to hereafter as a "programmed textbook." A programmed textbook is constructed with several items to a page, but with the sequence of items running from page to page rather than from top to bottom on the same page. That is, item 1 of lesson 1 appears at the top of the first page of the booklet and contains one or more blanks to be filled in. Item 2 appears at the top of page 2, item 3 at the top of page 3, and so on. The answer to item 1 is given at the top of page 2, just to the right of item 2, and the answer to item 2 is given at the top of page 3, just to the right of item 3, etc. In using such a booklet, the student reads item 1, writes his answer on a separate answer sheet, turns to page 2, checks his answer to item 1 against that in the upper right-hand corner of page 2; reads item 2, responds, turns to page 3, checks his answer to item 2 against that in the upper right-hand corner of page 3; and proceeds in like manner through each item in the lesson.

In the genetics program, there were 13 lessons with a total of 461 items, each of which had one or more blanks. These lessons were mimeographed and bound into four inexpensive booklets. An additional booklet of 4 lessons comprising 91 review items was also prepared. The several types of items (lead-in, augmenting, etc.) were employed with no particular emphasis on any one, except for the review lessons. Preliminary testing, conducted during the summer of 1959 with the assistance of several students, provided the information necessary to correct initial inadequacies of the program prior to its use. Analysis of responses from the 1959 experiments enabled the correction of additional inadequacies previous to the 1960 experiment.

In addition to the programmed textbook, a booklet described hereafter as the "experimental textbook" was also prepared. The experimental textbook contained the same material as the programmed textbook, for the most part being an identical, word-for-word reproduction of the programmed text. It was arranged in paragraph form and had no blanks to be filled in or problems to be solved except that at the end of each "chapter," items from the booklet of review lessons, described above, were included. It was thus much more like the conventional text which a student reads and studies in a customary manner.

These materials were employed in an experimental design conceived as a test of the relative effectiveness of learning with programmed materials in comparison with more conventional study and teaching techniques. In the calendar of the introductory course, the section on meiosis and basic genetics takes up one week of classes, i.e., three 1-hour lectures and one 3-hour laboratory. Thus the students were given one week in which to review mitosis and master the complexities of meiosis and basic genetics, whether they studied the programmed textbook, the experimental textbook, or neither.

Essentially the same experimental design was used at Oberlin and Earlham Colleges. Both institutions employ a single lecture section and multiple laboratory sections in the introductory biology course. In the 1959–60 experiment all students in five laboratory sections from each college were used: three sections, (the experimental groups), used the programmed materials; two sections served as control groups. In the 1960–61 experiment, six groups were used, three of these serving as controls. The treatment of each group was as follows:

Experimentals: (Oberlin and Earlham). These groups did not attend the three lectures presented during the experimental period.

Group 1—used the programmed textbook of 13 lessons, comprising 461 items each with one or more blanks to be filled in.

Group 2—used the same materials as group 1; in addition, this group used the booklet of 91 review items arranged into four review lessons. The review lessons occurred at the end of the 4th, 7th, 11th, and 13th lessons in the programmed textbook.

Group 3—used the experimental textbook exclusively.

Controls:

Group 4—(Oberlin 1959)—attended the three lectures and were strongly encouraged to study their textbook, *General Zoology,* (Villee, Walker, and Smith), an additional six hours; this was done by almost all of the students according to their time records.

Group 5 (Oberlin and Earlham)—attended the three lectures, studied their textbook (at Earlham the textbook was *Life* by Simpson, Pittendrigh, and Tiffany), but were not especially directed to study in any particular way or for any particular time. This group was told simply that they were being used as controls. In the 1960 Oberlin experiment, two sections constituted group 5. Tabulations were maintained separately on these two sub-groups to determine effects by class inasmuch as one group (5a) was composed of 42.5 percent freshmen, whereas group 5b was 77.5 percent freshmen.

Group 6 (Earlham and Oberlin, 1960)—did not attend the three lectures, but were assigned to independent study and referred to the same pages of their textbook as the members of group 5.

Pretest and post-test examinations were administered just prior to and at the conclusion of the experimental period. The pretest and post-test were different examinations. The same post-test was administed to the groups five months later in an attempt to measure retention.

RESULTS

Table 1 shows the mean pretest and post-test scores and the mean gain scores computed as the difference between pretest and post-test scores for each of the groups at Oberlin and Earlham. An analysis of covariance was applied to determine the significance of differences in means inasmuch as the groups were not initially uniform. In this analysis, each group was compared with every other

TABLE 1

Pretest and post-test mean scores and derived mean gain scores

Group	Number of students	Pretest score ($\bar{X}_1$)	Post-test score ($\bar{X}_2$)	Gain score ($\bar{X}_2-\bar{X}_1$)
Oberlin, 1959				
1	40	31.3	69.3	38.0
2	37	31.5	70.4	38.9
3	35	29.3	69.1	39.8
4	34	31.4	64.5	33.1
5	38	32.9	58.4	25.5
Earlham, 1959				
1	33	35.1	53.5	18.4
2	32	35.3	57.7	22.4
3	30	41.1	58.1	17.0
5	32	34.6	56.5	21.9
6	34	38.5	44.2	5.7
Oberlin, 1960				
1	37	37.7	54.8	17.1
2	41	42.3	57.7	15.4
3	37	41.0	54.0	13.0
5a	32	43.3	53.9	10.6
5b	28	45.2	51.6	6.4
6	38	49.5	49.5	0.0

TABLE 2

Results of the analysis of covariance

Groups	Oberlin, 1959	Earlham, 1959	Oberlin, 1960
1 vs. 2	NS*	NS	NS
1 vs. 3	NS	NS	NS
1 vs. 4	$.10>P>.05$	—	—
1 vs. 5	$P<.001$	NS	—
1 vs. 5a	—	—	NS
1 vs. 5b	—	—	NS
1 vs. 6	—	$P<.001$	$.01>P>.001$
2 vs. 3	NS	NS	NS
2 vs. 4	$.10>P>.05$	—	—
2 vs. 5	$P<.001$	NS	—
2 vs. 5a	—	—	NS
2 vs. 5b	—	—	$.05>P>.01$
2 vs. 6	—	$P<.001$	$P<.001$
3 vs. 4	$.10>P>.05$	—	—
3 vs. 5	$.01>P>.001$	NS	—
3 vs. 5a	—	—	NS
3 vs. 5b	—	—	NS
3 vs. 6	—	$P<.001$	$.05>P>.01$
4 vs. 5	$.10>P>.05$	—	—
5 vs. 6	—	$P<.001$	—
5a vs. 6	—	—	$.05>P>.01$
5b vs. 6	—	—	NS
5a vs. 5b	—	—	NS

*NS = Not statistically significant.

group, resulting in a set of 25 comparisons of Oberlin College groups (10 for 1960 and 15 for 1961) and 10 comparisons of Earlham groups. This procedure resulted in 35 F-ratios, each of which was evaluated for level of significance. The P value for each of these comparisons appears in table 2 and shows the following results:

Oberlin 1959: a) Groups 1, 2, and 3 are not significantly different from each other, but each is significantly different from Group 5; in addition the difference between each of these groups and Group 4 approaches significance at the .05 level of confidence.

b) The difference between groups 4 and 5 approaches significance at the .05 level of confidence.

Earlham: Groups 1, 2, 3, and 5 are not significantly different from each other, but each of these groups is significantly different from group 6 beyond the .001 level of confidence.

Oberlin 1960: a) Groups 1, 2 and 3 are not significantly different from each other but each is significantly different from group 6. Group 5a is also significantly different from group 6.

b) Group 2 is significantly different from group 5b.

TABLE 3
Time spent in study and improvement index

Group	N	Time (hr) ($\overline{X}$)	Improvement ($\overline{X}$)
Oberlin, 1959			
1	40	8.3	1.1
2	37	9.2	1.1
3	35	7.6	1.2
4	34	9.7	1.1
5	38	7.8	0.8
Earlham, 1959			
1	31	11.2	0.7
2	32	12.8	0.8
3	29	13.4	0.6
5	31	8.4	0.7
6	31	9.5	0.2
Oberlin, 1960			
1	37	8.1	0.8
2	41	7.8	0.6
3	37	8.4	0.7
5a	32	5.3	0.5
5b	28	5.8	0.3
6	38	5.5	0.3

In addition to the data on test scores, the mean number of hours of study spent by each group was computed from time records (groups 1 and 2) or estimates (groups 3 to 6) made by the participants. Also an "improvement index" was calculated for each student and expressed as "mean improvement" for each of the several groups. The improvement index is an arbitrary measure derived for each student by dividing the difference between the pretest and post-test scores by one-half the post-test score. The means of these parameters are shown in table 3. Inasmuch as there is unreliability in time estimates in groups 3 to 6, no correlations were sought between time spent in study and improvement score. The major results as to the effectiveness of programming might be contaminated by the variable of time spent but there did not appear to be any consistent trend through the samples.

DISCUSSION AND INTERPRETATION

From the data thus far presented, the following generalizations and interpretations seem warranted:

Oberlin, 1959: 1. The three groups using experimental materials (groups 1, 2, and 3) performed equally well and significantly better than group 5 which followed a conventional learning procedure. In addition the three experimental groups performed better than group 4 which used a conventional textbook for independent study; in each of these cases the difference between the means approaches significance.

2. The group 4 data present something of an enigma. In every comparison involving group 4, differences in performance approach, but do not quite reach, statistical significance. Group 4 did somewhat less well than any of the experimental groups and somewhat better than group 5, the "conventional" control. It is possible that this situation reflects a motivational variable inasmuch as group 4 was urged to make a special effort in studying the textbook materials; furthermore, they knew about the special study methods being used by the experimental groups. For the members of this group textbook study was somewhat more effective than that of group 5. This may well be an instance of the so-called "Hawthorne" effect, in which the mere knowledge that they are subjects in an experiment produces an increment in performance. Perhaps a more feasible explanation is simply that having been told to study the textbook an additional six hours served to emphasize the importance of the material on genetics and, perceiving this material as especially important, the students were more careful and perceptive in reading through it.

Earlham: 1. The three groups using experimental materials (groups 1, 2, and 3) appear to have done no worse and no better than group 5, which followed a conventional learning procedure, and significantly better than group 6 which used a conventional textbook for independent study.

2. Group 3 showed less improvement than either of the other two experimental groups, but it should be noted that gain scores for group 3 are computed from a higher pretest performance than either of the other two groups (1 and 2). In any event, the attained level of competence of this group was on par with the other two experimental groups and with group 5.

3. The gain score for group 6 is so low as to suggest that the group may have been atypical in some respect.

Oberlin, 1960: 1. The three groups using experimental materials performed as well and showed greater gain scores than groups 5a and 5b, both of which followed a conventional learning procedure, and significantly better than group 6, which used a conventional textbook for independent study.

2. Group 5b, which was 78 percent freshmen, showed less improvement, a lower post-test score, and a lower mean gain score than group 5a (43 percent freshmen). The score of this group (5b) is significantly different from group 2.

3. Group 6 which had the highest pretest score showed no mean gain and attained a lower post-test score than either group 5a or 5b.

Student Reaction

Student reaction to the experimental materials and procedures was sampled in the 1960 Oberlin experiment. In each experimental group the majority was favorably disposed toward the materials and procedures used (group 1—82 percent, group 2—65 percent, group 3—81 percent); the remaining students were about equally divided between neutral and unfavorable reactions. Of the students in group 6 only 41 percent reacted favorably to their procedure, 52 percent were neutral, and 7 percent were unfavorable to the use of a conventional textbook in independent study.

Retention Analysis

A retention examination, consisting of the same set of questions used for the post-test, was administered to each group five months after the post-test. The results in each case were biased owing to the fact that many students at both ends of the grade distribution, both on the post-test and in their course work, did not take the retention examination. Individual groups were thus biased in favor of or against the "more able" student.

The biases in each of the retention test samples do not permit conclusive statements on the relative retention of programmed learning of the type used in these experiments as compared with retention of conventionally-learned materials. It does appear, however, that there is no significant difference in retention and that the use of such programs can neither be justified nor negated on this basis. The factor of relative retention may have bearing upon the decision to use or not to use programming techniques and teaching machines. Certainly this aspect of programmed learning needs more attention than has been so far given.

SUMMARY AND CONCLUSIONS

1. Students appear to learn more effectively and efficiently in independent study with materials designed for independent study than with conventional textbooks. The conventional textbook is more frequently designed to complement and supplement a traditional learning program.
2. There is no statistically significant difference between the groups using the programmed textbook procedure with and without reviews; it may be concluded that additional review is unnecessary since it adds nothing to the group using it.
3. The groups using either textbook method (programmed or experimental) appear to have done equally well; therefore either might be used for independent study, depending upon the preference of the student or instructor.
4. There are no data in this experiment which suggest that the programmed textbook is superior as a study technique to the experimental textbook. However, there is strong evidence that for independent study either of these methods is superior to "conventional" instructional techniques and result in a better level of learning. The explanation of this difference may simply reflect the fact that the experimental materials were very carefully organized and designed to be as clear and unambiguous as possible. The experimental materials were programmed first and then written up in the experimental textbook form; thus, the organization of content produced by the demands of programming procedures inevitably carried over into the experimental textbook content.

LITERATURE CITED

Crowder, N. A. 1960. Automatic tutoring by intrinsic programming, p. 286–298. *In* A. A. Lumsdaine and R. Glaser, [ed.], Teaching machines and programmed learning. Natl. Educ. Assoc., Washington.

Galanter, E., [ed.]. 1959. Automatic teaching: the state of the art. John Wiley and Sons, Inc., New York. 198 p.

Lumsdaine, A. A., and **R. Glaser,** [ed.]. 1960. Teaching machines and programmed learning. Natl. Educ. Assoc., Washington. 724 p.

Pressey, S. L. 1926. A simple apparatus which gives tests and scores—and teaches. School and Soc. 23: 373–376.

———. 1960. Some perspectives and major problems regarding "teaching machines," p. 497–505. *In* A. A. Lumsdaine and R. Glaser, [ed.], 1960. Teaching machines and programmed learning. Natl. Educ. Assoc., Washington.

Silverman, R. E. 1960. Automatic teaching: a review of theory and research. U. S. Naval Training Device Center, Port Washington, N. Y. 47 p.

Skinner, B. F. 1958. Teaching machines. Science 128: 969–977.